野球試合映像からの高速な全投球映像抽出手法の検討

足立 翔平† 福田 大翔† 渡辺 裕†

†早稲田大学大学院 基幹理工学研究科 情報理工・情報通信専攻
〒169-0072 東京都新宿区大久保 3-14-9 早大シルマンホール 401 号室
E-mail: †{alice-fr@asagi.waseda.jp, taketomohiro@akane.waseda.jp, hiroshi.watanabe@waseda.jp}

あらまし TV のスポーツ番組や報道番組において，野球の試合終了後にハイライトとして全投球シーンを抽出し，放送する需要が存在する．こうした全投球映像を作成するためには，映像編集に携わる労働コストが必要となる．加えて，試合終了後速やかに全投球映像を放送するために，映像編集作業の速度も要求される．そこで，本稿では，姿勢推定技術及び画像分類技術を用いた，半自動且つ高速な全投球映像抽出手法を検討する．

キーワード 野球，投球，姿勢推定，画像分類

A Study on Fast Extraction Method of All Pitches from Baseball Game Video

Shohei ADACHI† Hiroto FUKUTA† and Hiroshi WATANABE†

†Graduate School of Fundamental Science and Engineering, Waseda University
#401 Sodai Shillman hall Building 3-14-9 Okubo, Shinjuku-ku, Tokyo, 169-0072 Japan
E-mail: †{alice-fr@asagi.waseda.jp, taketomohiro@akane.waseda.jp, hiroshi.watanabe@waseda.jp}

Abstract In TV sports and news programs, there is a demand for extracting and broadcasting all pitching scene as a highlight after a baseball game. In order to create such all-pitch video, labor costs for video editing are required. In addition, the speed of video editing work is also required to broadcast all pitching scenes immediately after the game. In this paper, we propose a semi-automatic and fast extraction method of all pitching videos using pose estimation and image classification techniques.

Keyword Baseball, Pitching, Pose Estimation, Image Classification

1. はじめに

TV のスポーツ番組や報道番組において，試合のハイライトとなるシーンを放送する需要が存在する．野球においては，投球シーンがハイライトとして扱われるが，例としてパ・リーグの一試合における投球数は約 300 球という膨大な数である[1]．そのため，全投球シーンを抽出し，ハイライト映像を作成するための映像編集には多大な労働コストが必要となる．また，試合後の速やかな放送のためには映像編集作業の速度も要求される．上述した労働コストや高速な編集作業の必要性という課題を解決するため，本稿では姿勢推定技術及び画像分類技術を用いた，半自動且つ高速な野球試合映像からの全投球映像抽出手法を検討する．また，分類の際に扱う姿勢情報として，骨格画像と骨格座標値の二つの条件を検討する．

2. 関連技術

2.1. YOLOv8

YOLO[2]は画像を入力とし，CNN を利用した高速な物体検出を可能とするアルゴリズムであり，様々なバージョンやモデルが存在する．本研究においては投球動作の認識に用いるための姿勢情報を得る目的で，最新の YOLOv8[3]における姿勢推定モデルであり，骨格情報をリアルタイムで出力可能な YOLOv8n-pose を利用する．姿勢推定結果の骨格画像を図 1 に示す．

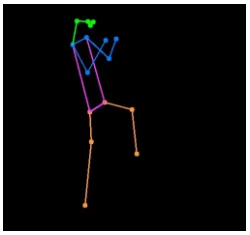

図 1 骨格推定画像（YOLOv8）

2.2. ResNet18

ResNet[4]は CNN を利用した画像認識モデルである．勾配消失問題を解決しつつ CNN の層を深くするため，Residual Block の形式がとられている．ResNet には用いられる CNN の層数によって複数のモデルが存在するが，本研究においては処理量の観点から最小の層数を持つ ResNet18 を用い，これをファインチューニングすることで投球動作の認識に利用する．

3. 提案手法

3.1. 概要

骨格情報を用いた投球シーン切り抜き手法の概要を図 2 に示す.

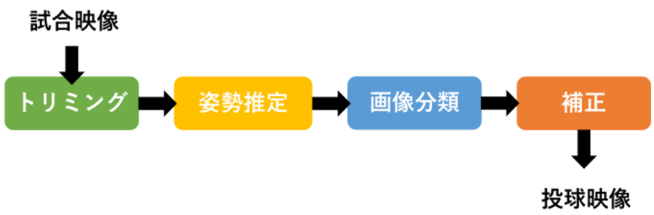

図 2 提案手法 概要

試合映像を入力し，投球動作を行う投手の周囲でトリミングする．このトリミングは，映像の画角や投手の画面に占める割合の差異による影響を抑え，多様な入力映像に対して分類性能を頑健にすることを目的としている．本研究においては手動でトリミングの領域を設定しているが，YOLOv8 を投手の画像でファインチューニングすることにより，自動的なトリミング領域の設定も可能となる.

続いて，トリミング画像に対して姿勢推定を行う．画像分類において骨格情報のみを扱うことで，服装や背景，画面に映っている他の人物の影響を抑制する.

最後に，分類結果に補正や閾値の設定を施し，投球と判定されたフレーム群を動画として出力することで，投球映像の切り抜きを可能とする.

3.2. 姿勢推定

姿勢推定には 2.1 節で述べた YOLOv8n-pose を用いる．推定結果は骨格画像として出力する場合と，骨格座標値のみを出力する場合の 2 条件を検討する.

3.2.1. 骨格画像の出力

YOLOv8n-pose による推定結果画像は通常，入力画像上に重なる形で骨格画像が表示される．しかし，映像上の人物の服装や，背景の情報が画像分類に与える影響を抑えるため，入力画像情報を 0 の値で置換し黒色画像にすることで，骨格のみを表示する.

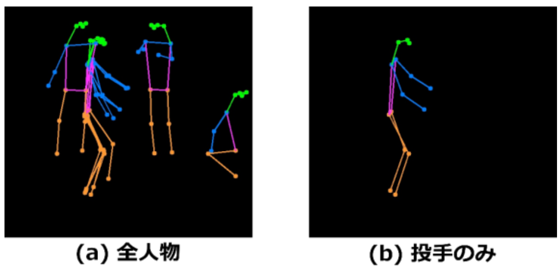

図 3 骨格画像例

また，図 3(a)に示すように，通常の出力では画像内で検出された全ての人物の骨格が表示される．この場合，注目すべき投手の他に，打者や捕手，観客の骨格情報も表示されてしまう．加えて，図 3(a)における投手の骨格に注目すると，一人の人物に対して複数の骨格が重なって表示されている．映像のブラーの影響から，一人の人物に対して複数回検出された際にこのような表示が発生するが，これも画像分類に悪影響を及ぼす原因となる．そこで，投手周辺をトリミングした場合，カメラと人物との距離の関係上，画像内の投手の位置が最も下部に来ることから，検出された人物の中で最下部に位置する人物を投手とし，投手の骨格情報一つのみを骨格画像として出力することとする．この処理の結果得られる骨格画像の例を図 3(b)に示す.

3.2.2. 骨格座標値の出力

骨格画像は黒色画像の背景上に骨格を結ぶ線という非常にシンプルな情報でありながら，入力映像のトリミング画像と同サイズの画素数を CNN に入力するため，分類に要する時間効率が悪いと考えられる．そこで，骨格を形成する情報をテンソル形式の骨格座標値という小さいデータサイズの形式で CNN に入力することで，分類処理のパラメータ削減を試みる.

この際，座標値は入力した画像サイズによって，同じ姿勢･同じ位置の骨格であっても異なる数値を取る．そこで，x，y 座標値をそれぞれ入力画像の幅，高さで正規化することで 0~1 の範囲の値とし，入力画像サイズによる影響を排除する.

3.3. 画像分類

2.2 節で述べた ResNet18 を投球データでファインチューニングして用い，トリミング画像から推定した骨格画像を入力として画像を分類する．この際，分類は 3 クラスとし，その他，投球動作，予備・予後動作の三つに分類する．また，分類結果のフラグを 0：その他，1：投球動作，2：予備・予後動作として設定し，入力映像のフレーム数に対応した配列要素に分類結果のフラグを出力する．投球動作が一回含まれる映像を入力とした際の分類結果の例を図 4 に示す．動作の遷移の様子が確認できる.

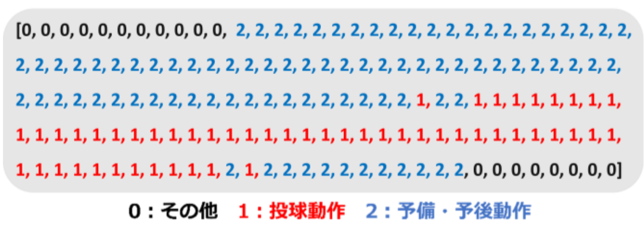

図 4 画像分類結果例

また，骨格座標値はテンソル形式でモデルに入力し，小さいサイズの画像として扱う．この際，骨格画像と比較して非常に小さなデータサイズであるため，学習モデルは ResNet18 を用いず，3 層の CNN と 1 層の全結合層からなるネットワークを用いる.

3.4. 分類結果の補正・出力

画像分類結果を基に，1：投球動作と分類されたフレームのみを映像として出力することで，投球動作映像の切り抜きが完了する．しかし，図 5 において動作の切り替わり部分に注目すると，1,2,2,1 や 1,1,2,1 のように投球フレームが不連続になる部分が存在する．これは，画像分類における誤検出が原因である．この分類結果のまま投球動作フレームのみを映像として出力すると，映像の開始・終了付近で動作が不連続となる．そこで，分類結果の不連続性を除去する補正処理を行った後，映像として出力することとする．

補正方法として，分類結果を格納した配列を最初の要素から N 要素ずつのグループ毎に調査し，この N 要素の中に一つでも 1：投球動作と分類されたものがあった場合，N 要素全てを 1：投球動作に置換する手法を用いる．N=3 の場合の補正例を図 5 に示す．

図 5 画像分類結果補正例

また，牽制球や個々人の投球前の習慣的行為であるルーティン動作を投球として出力することを防ぐため，出力段階で閾値を設定可能にする．これは，牽制球やルーティン動作が投球動作と比較して短い動作となる傾向に注目している．設定した場合，1：投球動作と判定されたフレームが閾値以上連続した場合のみ，そのフレーム群を投球映像として出力する．

4. 実験

4.1. データセット

トレーニング用のデータセットとして，投球動作のフレーム約 8000 枚，予備・予後動作のフレーム約 2400 枚，その他のフレーム約 2400 枚を用いる．なお，これらのデータはトレーニングするモデルに応じて，元画像，元画像＋トリミング，骨格画像，骨格画像＋トリミング，骨格座標値の 5 種類を用意する．

また，テスト用のデータセットとして，トレーニングデータとドメインが類似した試合映像（類似ドメイン映像），ルーティン動作を含む類似ドメイン映像，非類似ドメイン映像の三つの映像を用いる．表 1 にテストデータセットの詳細を示す．

表 1 テストデータセット

	類似ドメイン	類似ドメイン（ルーティン含む）	非類似ドメイン
サイズ [px]	1920×1080	1920×1080	1920×1080
時間 [min]	25	25	19.5
フレームレート [fps]	30	30	30
投球数	55	43	42

4.2. 評価指標

本実験の評価指標として，適合率，再現率，F 値を用いる．適合率は，本実験において投球として切り抜かれた映像の内，実際に投球であるものの割合を示す．再現率は，本実験におけるテスト映像の全投球の内，実際に切り抜かれた投球の割合を示す．F 値は適合率と再現率のトレードオフを示し，両者の調和平均で表される．また，処理時間として試合映像を入力してからトリミング，姿勢推定，画像分類，補正を経て投球映像を出力するまでにかかる時間を計測する．

4.3. 実験結果

実験結果を表 2,3,4 に示す．なお，各条件の比較のため，“元画像”，“元画像＋トリミング”，“元画像＋トリミング＋閾値”，“骨格画像”，“骨格画像＋トリミング”，“骨格画像＋トリミング＋閾値”，“骨格座標値”の計 7 種の手法の結果を示す．

表 2 各条件における実験結果（1）

手法	類似ドメイン			
	適合率	再現率	F値	時間 [sec]
元画像	0.856	0.867	0.861	335
元画像＋トリミング	0.922	0.994	0.957	**103**
元画像＋トリミング＋閾値	**1.000**	0.939	0.969	107
骨格画像	0.926	0.976	0.950	536
骨格画像＋トリミング	0.970	0.976	**0.973**	169
骨格画像＋トリミング＋閾値	**1.000**	0.867	0.929	164
骨格座標値	0.829	**1.000**	0.907	128

表 3 各条件における実験結果（2）

手法	類似ドメイン（ルーティン含む）			
	適合率	再現率	F値	時間 [sec]
元画像	0.683	0.853	0.759	339
元画像＋トリミング	0.741	0.992	0.848	114
元画像＋トリミング＋閾値	0.869	0.922	0.895	**99**
骨格画像	0.955	0.984	0.970	555
骨格画像＋トリミング	0.977	**1.000**	**0.988**	167
骨格画像＋トリミング＋閾値	**0.981**	0.845	0.908	161
骨格座標値	0.781	0.992	0.874	128

表 4 各条件における実験結果（3）

手法	非類似ドメイン			
	適合率	再現率	F値	時間 [sec]
元画像	0.868	0.159	0.268	270
元画像＋トリミング	**1.000**	0.738	0.849	**118**
元画像＋トリミング＋閾値	**1.000**	0.548	0.708	**118**
骨格画像	**1.000**	0.762	**0.865**	439
骨格画像＋トリミング	0.926	0.722	0.811	165
骨格画像＋トリミング＋閾値	0.947	0.516	0.668	165
骨格座標値	0.634	**0.992**	0.773	136

表 2,3,4 に示す実験結果について，トリミングを加えることで適合率及び再現率の向上が確認できる．トリミングによって，動作を行う人物の情報を重点的に扱うことが可能になるためと考えられる．

閾値を加えた場合は適合率の向上が見られる一方，再現率は低下している．ルーティンや牽制球の動作が投球よりも短い時間で行われる点に注目して閾値を設けたため，適合率の向上には貢献する一方で，素早い投球動作も同様に省いてしまった結果と考えられる．

同じ条件下での“元画像”と“骨格画像”の結果を比較した場合，適合率及びF値は多くの場合で“骨格画像”の性能が高い．一方で，再現率は原寸サイズ以外では“元画像”の性能が高い．これは，元画像は背景や色などの多くの情報を持つため，投球と認識可能な範囲が広く，再現率が高くなる一方で，骨格画像は骨格情報のみなため，投球と認識するフレームの確度が高く，結果として適合率が高くなったと考えられる．

処理時間に関しては，同条件で比較した場合，元画像のまま扱う方が高速となっている．これは，骨格画像を扱う際は姿勢推定の処理が加わっている点に起因する．また，トリミングの有無で比較した場合，トリミングした場合の方が2倍以上高速であり，特に“骨格画像”の場合に顕著となっている．これは，トリミングによって映像からフレームへの変換や，画像分類の際に扱う情報量が削減されるためであり，特に“骨格画像”については姿勢推定の処理量も削減できるため，より効果的となっている．

“骨格座標値”の適合率及びF値は“元画像＋トリミング”や“骨格画像＋トリミング”より低い値だが，再現率については他2手法よりも高く，非類似ドメインについては突出している．これは，“元画像”の再現率が高くなった理由と対称的に，扱う情報量が少ないために投球動作及びそれに近しい動作を全て投球と判定した結果，適合率が大きく低下し，その分再現率が高くなったと考える．また，処理時間は“元画像＋トリミング”より遅いものの，“骨格画像＋トリミング”より高速となっている．これは，骨格推定処理が加わっているため元画像のまま扱うよりは遅いものの，分類の際に骨格画像よりも小さな情報量を扱っているためネットワークのパラメータが小さく，結果として両者の間の実行時間になったといえる．

5. 考察

総合的な評価のために，七つの手法における3種のテストデータの結果の平均値を表5に示す．

表5 各条件における実験結果（平均値）

手法	適合率	再現率	F値	時間 [sec]
元画像	0.802	0.626	0.629	315
元画像＋トリミング	0.888	0.908	0.885	112
元画像＋トリミング＋閾値	0.956	0.803	0.857	**108**
骨格画像	0.960	0.907	**0.928**	510
骨格画像＋トリミング	0.958	0.899	0.924	167
骨格画像＋トリミング＋閾値	**0.976**	0.743	0.835	163
骨格座標値	0.748	**0.995**	0.851	131

表8の結果から，適合率と再現率のバランスであるF値において高い性能を示したのは“骨格画像”及び“骨格画像＋トリミング”となった．両者のF値には約0.4%の差異しかない一方，処理時間について比較すると“骨格画像＋トリミング”の方が約70%短いことが確認できる．そのため，七つの手法において総合的な性能が最も優れているのは“骨格画像＋トリミング”と考える．また，再現率について着目した場合，“骨格座標値”が突出して高い性能を示している．そのため，不足なく全投球を切り抜くことを重視する場合は，“骨格座標値”を利用することが有効といえる．

また，実際の試合時間・投球数に近づけるため，類似ドメインのテスト映像を6倍し，試合時間が150分，投球数が330球となった場合を仮定する．この場合，“骨格画像＋トリミング”は約17分で322球の投球映像と10個の誤検出映像を出力する．“骨格座標値”は約13分で全投球映像と68個の誤検出映像を出力する．上記計算結果はトリミングの領域選択に要する時間も6回分で計算しているが，実際にはトリミングの領域選択は一度のみ行われるため，計算上の処理時間よりも短縮される．

6. まとめ

本稿では，野球の試合映像から全投球映像を抽出する編集作業にかかる労働コストの削減及び編集作業の高速化を目的として，姿勢推定技術と画像分類技術を利用した半自動且つ高速な野球試合映像からの全投球映像抽出手法を提案し，七つの条件における実験評価及び考察をした．結果として，トリミングした骨格画像を扱う手法が総合的に最も高い性能を示し，不足なく全投球映像を切り抜くという目的の下においては骨格座標値を扱う手法に優位性があることを示した．

今後の展望として，物体検出モデルによるトリミング領域の自動選択の実装や，分類モデルの学習時に前後のフレーム情報を考慮させることによる，更なる動作分類精度の向上が考えられる．

文　献

[1] パ・リーグ インサイト，“総投球数は12万球超え。パ・リーグにおける今季の球種割合を調査してみた”，（最終閲覧日：2024年2月20日 https://pacificleague.com/news/56208

[2] J. Redmon, S. Divvala, R. Girshick and A Farhadi, “You Only Look Once: Unified Real-Time Object Detection,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), pp. 779-788, Jun. 2016.

[3] Ultralytics/Ultralytics,[online]Available:https://github.com/ultralytics/ultralytics.

[4] K. He, X, Zhang, S. Ren and J Sun, “Deep Residual Learning for Image Recognition,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), pp. 770-778, Jun. 2016.